MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Deep Graph Topology Learning for 3D Point Cloud Reconstruction

Duan, Chaojing; Chen, Siheng; Tian, Dong; Moura, Jose; Kovacevic, Jelena

TR2019-046 June 29, 2019

Abstract

We propose an autoencoder with graph topology learning to learn compact representations of 3D point clouds in an unsupervised manner. As discrete representations of continuous surfaces, 3D point clouds are either directly acquired via 3D scanners like Lidar sensors, or generated from multi-view images or RGB-D data. Different from 1D speech data or 2D images, which are associated with regular lattices, 3D point clouds are usually sparsely and irregularly scattered in the 3D space; this makes traditional latticed-based algorithms difficult to handle 3D point clouds. Most previous works discretize 3D point clouds by transforming them to either 3D voxels or multi-view images, causing volume redundancies and the quantization artifacts. As a pioneering work, PointNet is a deep-neural-network based method that uses pointwise multi-layer perceptron followed by maximum pooling to handle raw 3D points and achieve remarkable performances in many supervised tasks, including classification, segmentation and semantic segmentation of 3D point clouds.

Graph Signal Processing Workshop (GSP)

Deep Graph Topology Learning for 3D Point Cloud Reconstruction

Chaojing Duan[1], Siheng Chen[2], Dong Tian[3], José M. F. Moura[1], Jelena Kovačević[4]

[1] Carnegie Mellon University, [2] Mitsubishi Electric Research Laboratories (MERL),
[3] InterDigital, [4] New York University

We propose an autoencoder with graph topology learning to learn compact representations of 3D point clouds in an unsupervised manner. As discrete representations of continuous surfaces, 3D point clouds are either directly acquired via 3D scanners like Lidar sensors [5], or generated from multi-view images or RGB-D data [4]. Different from 1D speech data or 2D images, which are associated with regular lattices [3], 3D point clouds are usually sparsely and irregularly scattered in the 3D space; this makes traditional latticed-based algorithms difficult to handle 3D point clouds. Most previous works discretize 3D point clouds by transforming them to either 3D voxels or multi-view images, causing volume redundancies and the quantization artifacts. As a pioneering work, PointNet is a deep-neural-network based method that uses pointwise multi-layer perceptron followed by maximum pooling to handle raw 3D points and achieve remarkable performances in many supervised tasks, including classification, segmentation and semantic segmentation of 3D point clouds [6].

In this work, we consider unsupervised learning of 3D point clouds; that is, learning compact representations of 3D point clouds via self-organization. Several recent works have been proposed to pursue a similar goal, such as LatentGAN [1], 3DGAN [7], and FoldingNet [8]. They adopt an encoder-decoder framework. The encoder follows similar architectures in PointNet and extracts global features; and the decoder is used to reconstruct 3D point clouds based on global features produced by the encoder. To design a decoder, LatentGAN [1] uses fully-connected layers, which does not explore the geometric properties of 3D point clouds at all; FoldingNet [8] uses point-wise multi-layer perceptrons to fold a 2D lattice to a 3D surface, which assumes that the underlying surface of all points has genus less than 2. Table 1 shows that FoldingNet can hardly reconstruct torus with high-order genus. Moreover, since features obtained from the encoder provide global information, previous works are hard to capture detailed local geometric structures.

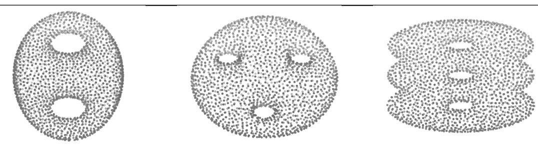
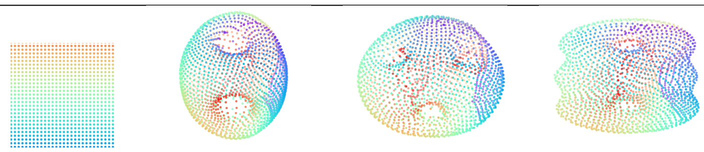
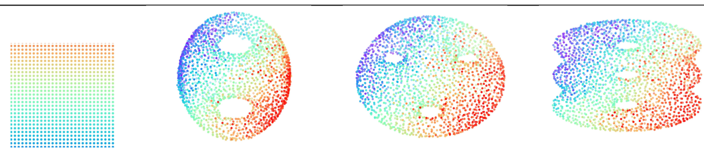
Input	
FoldingNet *without graph filtering*	
FoldingNet++ *with graph filtering*	

Table 1: **Graph filtering refines the reconstructions of 3D point clouds.** The first row shows the original point clouds sampled from torus generated in MeshLab [2]; the second row shows the coarse reconstructions obtained by the folding module; and the third row shows the refined reconstructions obtained after graph filtering. The color associated with each point indicates the correspondence between a node in the 2D lattice and a 3D point. The smoothness of the color transition reflects the difficulty of the folding process. Only training the folding module cannot capture the holes and the folding process is difficult. With the learn graph structures, the networks achieve much finer reconstructions.

To solve this issue, we extend FoldingNet by proposing a novel graph-topology-learning module, which guides the networks to handle complex shapes and explore local geometric structures. We call the overall networks *FoldingNet++*. The encoder in FoldingNet++ still adopts PointNet and the decoder consists of three modules: the folding module and the graph-topology-learning module and the graph-filtering module. The folding module folds a regular 2D lattice to a coarse 3D point cloud and has similar architecture with FoldingNet. The graph-topology-learning module learns a graph shift operator to explicitly capture the pairwise relationships between 3D points. Intuitively, the graph shift operator is able to deform a 3D point cloud by *cutting* or *glueing* shapes locally. In other words, the folding module (FoldingNet) only reconstructs 3D positions, while the graph-topology-learning module pushes the networks to reconstructs both 3D positions and pairwise relationships for all the points. In the graph-filtering module, we finally design graph filters based on the learnt graph shift operator and filter the coarse 3D point cloud to obtain a refined 3D point cloud. We validate the effectiveness of FoldingNet++ in two tasks, including 3D point cloud reconstruction and transfer classification. The experimental results show that FoldingNet++ outperforms LatentGAN, 3DGAN and FoldingNet in all the tasks; see Table 2.

Method	Modality	# code	ModelNet10	ModelNet40
3D-GAN [7]	Voxels	7168	83.30%	91.00%
Latent-GAN [1]	Points	512	85.70%	95.40%
FoldingNet [8]	Points	512	88.40%	94.40%
FoldingNet++	Points	512	**89.67%**	**95.63%**

Table 2: **FoldingNet++ achieves best classification accurracies in two datasets**.

The main contributions of this work are as follows:

- We propose a novel deep-neural-network-based autoencoder, called FoldingNet++, to do unsupervised learning for raw 3D point clouds. FoldingNet++ uses a graph-topology-learning module to guide the networks to handle complex shapes and explore local geometric structures;
- We validate the proposed FoldingNet++ in the tasks of 3D point cloud reconstruction and transfer classification. The experimental results show that FoldingNet++ outperforms 3D-GAN, Latent-GAN and FoldingNet;
- We propose a 3D shape dataset with subcategory labels based on ModelNet40 and illustrate that the graph-topology-learning module guides FoldingNet++ to refine details.

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Representation learning and adversarial generation of 3d point clouds. *arXiv preprint arXiv:1707.02392*, 2017.

[2] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136, 2008.

[3] Chaojing Duan, Siheng Chen, and Jelena Kovačević. Weighted multi-projection: 3d point cloud denoising with tangent planes. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 725–729. IEEE, 2018.

[4] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2010.

[5] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics*, 43(5):1318–1334, 2013.

[6] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[7] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems*, pages 82–90, 2016.

[8] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 3, 2018.